HUMAN CARBOHYDRATE-ASSOCIATED PROTEINS

TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of human carbohydrate-associated proteins and to the use of these sequences in the diagnosis, treatment, and prevention of autoimmune/ inflammatory disorders, gastrointestinal disorders, infectious disorders, reproductive disorders, neurological disorders, eye disorders, and cell proliferative disorders, including cancer.

BACKGROUND OF THE INVENTION

Proteins are associated with carbohydrates in several ways. Glycoproteins have covalently attached carbohydrates which have been added to the proteins as they traverse the secretory pathway. Some proteins noncovalently associate with carbohydrate-containing macromolecules for purposes of binding, modifying, or degrading the carbohydrates. Carbohydrate-containing macromolecules, which include glycoproteins, glycolipids, glycosaminoglycans, and proteoglycans, are found on the cell surface and in the extracellular matrix. Glycosaminoglycans (GAGs) are linear unbranched polysaccharides composed of repetitive disaccharide units. GAGs exist free or as part of proteoglycans, large molecules composed of a core protein attached to one or more GAGs. (See Lodish, H. et al. (1995) Molecular Cell Biology, Scientific American Books, New York NY, pp. 1136-1142.)

Cell surface carbohydrate-containing macromolecules, including glycoproteins, glycolipids, and transmembrane proteoglycans, mediate adhesion with other cells and with components of the extracellular matrix. The extracellular matrix is comprised of diverse glycoproteins, GAGs, and carbohydrate-binding proteins which are secreted from the cell and assembled into an organized meshwork in close association with the cell surface. The interaction of the cell with the surrounding matrix profoundly influences cell shape, strength, flexibility, motility, and adhesion. These dynamic properties are intimately associated with signal transduction pathways controlling cell proliferation and differentiation, tissue construction, and embryonic development.

The GAG hyaluronan (HA) is found in the extracellular matrix of many cells, especially in soft connective tissues, and is abundant in synovial fluid. (Pitsillides, A.A. et al. (1993) Int. J. Exp. Pathol. 74:27-34.) HA seems to play important roles in cell regulation, development, and differentiation (Laurent, T.C. and Fraser, J.R. (1992) FASEB J. 6:2397-2404). HA functions in water and plasma protein homeostasis (Laurent, supra). HA may play a role in mitosis as its production increases in proliferating cells. HA is known to induce tumor cell adhesion and migration, and its small fragments are angiogenic. HA has been chemically modified as a

biomaterial for medical applications such as controlled drug release matrices, nerve guides, and wound dressings (Zhong, S.P. et al. (1994) Biomaterials 15:359-365). HA's tissue protective and rheological properties have proved useful in ophthalmic surgery. Serum HA is diagnostic of liver disease and various inflammatory conditions, such as rheumatoid arthritis. Interstitial edema caused by accumulation of HA may cause dysfunction in various organs (Laurent, supra).

Hyaluronidase is an enzyme that degrades HA to oligosaccharides by catalyzing the random hydrolysis of 1,4-linkages between N-acetyl-beta-D-glucosamine and D-glucuronate residues. Hyaluronidases have been found in organisms as diverse as bacteria, insects, reptiles, and mammals, and may function in cell adhesion, infection, angiogenesis, and signal transduction.

PH-20, a protein expressed in the mammalian testis and present on the plasma membrane of mouse and human sperm, has hyaluronidase activity (Lin, Y. et al. (1994) J. Cell Biol. 125:1157-63). PH-20 enables sperm to penetrate the mammalian egg's outer layer, which consists of about 3,000 cumulus cells embedded in an extracellular matrix rich in HA. Penetration of this layer is an essential step in the fertilization process. PH-20 is also expressed in some tumor cells. Non-testicular mammalian hyaluronidases also exist. The HYAL1 hyaluronidase is expressed in human serum while the lysosomal hyaluronidase HYAL2 is expressed in many cells (Lepperdinger, G. et al. (1998) J. Biol. Chem. 273:22466-22470). HYAL2 may have a role in producing distinct HA fragments that can induce angiogenesis and the expression of enzymes involved in signal transduction pathways, such as nitric oxide synthase. A lysosomal-type hyaluronidase may degrade HA in lung fibroblasts in a cytokine-regulated process (Sampson, P.M. et al. (1992) J. Clin. Invest. 90:1492-1503).

The venom of numerous animals including various snakes, bees, hornets, stone fish, platypus, scorpions, and lizards contain hyaluronidase. Hyaluronidase from the white face hornet is an allergen which induces an IgE response in susceptible people (Lu, G. et al. (1995) J. Biol. Chem. 270:4457-4465). Venom hyaluronidase is thought to act as a "spreading factor", an aid in the diffusion of toxins. Researchers have found that lizard venom hyaluronidase promotes the spread of the hemorrhagic area in mice injected with hemorrhagic toxin (Tu, A.T. and Hendon, R.R. (1983) Comp. Biochem. Physiol. B 76:377-383). Clostridium perfringens hyaluronidase is important for the pathogenicity of this bacterium, which causes food poisoning and gas gangrene (Canard, B. et al. (1994) Mol. Gen. Genet. 243:215-224). The hyaluronidase destroys connective tissue in the host, allowing Clostridium perfringens to spread from the initial site of infection and to colonize and attack surrounding tissue.

Hyaluronidases are associated with reproduction, cancer, and inflammation. Effective contraception (100%) was obtained in male and female guinea pigs immunized with PH-20

(Primakoff, P. et al. (1988) Nature 335:543-546). Based on these results researchers are continuing efforts to make an anti-sperm PH-20 contraceptive vaccine suitable for humans (Zhu, X. and Naz, R.K. (1994) Arch. Androl. 33:141-144). Hyaluronidase activity is significantly elevated in prostate tumor tissue compared to that in both normal prostate and benign prostate hyperplasia (Lokeshwar, V.B. et al. (1996) Cancer Res. 56:651-657). Furthermore, hyaluronidase levels in tissues correlate well with tumor progression. The combined therapy of hyaluronidase, vindesine, cisplatin, and radiation is highly effective against advanced squamous cell cancer of the head and neck and is well tolerated by most human patients (Klocker, J. et al. (1995) Am. J. Clin. Oncol. 18:425-428). Hyaluronidase in combination with the chemotherapeutic drug vinblastine had significant antitumor effects on SK-Mel-3 melanoma cells implanted in nude mice (Spruss, T. et al. (1995) J. Cancer Res. Clin. Oncol. 121:193-202). Furthermore, hyaluronidase was well tolerated in test animals and prevented the local inflammation reactions that are commonly seen after subcutaneous vinblastine injections.

N-Acetylglucosamine Receptor

Many secreted proteins and integral membrane proteins have oligosaccharide chains attached by *O*- or *N*-glycosidic linkage. *N*-linked oligosaccharides contain a common pentasaccharide consisting of three mannose and two N-acetylglucosamine residues. Additional sugars are often attached to this common backbone, either as additional mannose residues or as a complex array of N-acetylglucosamine, galactose, sialic acid, and L-fucose. A large number of surface sugar patterns are possible, thereby providing a large amount of information that may be of functional importance.

Lectins are carbohydrate-binding proteins, originally described in plants, that interpret the information provided by oligosaccharide moieties. The liver cell asialoglycoprotein receptor is a lectin that recognizes sialic acid residues. Immunoglobulins and peptide hormones, which contain terminal sialic acid residues, are removed from the blood after they bind to the liver receptor and are transported into liver cells by endocytosis. Selectins are cell surface receptors on neutrophils and other leukocytes that bind carbohydrate moieties on lymph node tissue, endothelium, and activated platelets. The selectins target leukocytes to sites of injury during an inflammatory response. Cell surface carbohydrates also appear to act as guides in neuronal outgrowth.

N-acetylglucosamine (GlcNAc)-binding proteins have been described in several cell types and are involved in a variety of functions. The chicken hepatic lectin (CHL) appears to be the avian homolog of the mammalian asialoglycoprotein receptor, but with a greater affinity for GlcNAc instead of sialic acid. CHL is a homotrimer in the plasma membrane of chicken liver cells and it

mediates the clearance of serum glycoproteins that have a terminal GlcNAc residue. The ability of ligand to bind to CHL requires Ca^{2+}, which is also the case with other Ca^{2+}-dependent (C-type) lectins such as asialoglycoprotein, mannose-binding protein (MBP), and selectin cell adhesion molecules. (Burrows, L. et al (1997) Biochem. J. 324:673-680.)

GlcNAc-binding proteins have been identified in the nucleus of human myeloid leukemia cells. These lectins bind to *O*-linked GlcNAc, present on nuclear glycoproteins. Interaction between lectins and nuclear glycoproteins inhibit trafficking of molecules between the nucleus and cytoplasm, and inhibit transcriptional activation by the transcription factor Sp1. (Felin, M. et al (1994) J. Cell. Biochem. 56:527-535.)

GlcNAc-binding protein also regulates trafficking of thyroglobin in the thyroid. Thyroglobin, the precursor of the thyroid hormones, is stored in the thyroid follicles as a prohormone. Production of the prohormone includes addition of a carbohydrate moiety as it passes through the ER and Golgi, and iodination of tyrosine residues in exocytic vesicles. Complete maturation of prohormone takes 24-36 hours and requires successive recycling through the iodination steps. Stimulation of the thyroid with thyroid stimulating hormone (TSH) causes follicular cells to take up thyroglobin from the follicular lumen by pinocytosis. The prohormone moves into an endosomal compartment, where immature thyroglobin is picked up by an endosomal receptor and passed back to the thyroperoxidase iodination site via the Golgi. Only mature prohormone is passed to the lysosome for degradation, which produces the thyroid hormones thyroxine (T_4) and triiodothyronine (T_3). The thyroglobin domain on the immature prohormone interacts with the endosomal receptor through N-acetylglucosamine (GlcNAc) residues. The domain also includes tyrosyl residues involved in iodination and hormonogenesis. The thyroglobin receptor is an endogenous lectin showing specificity for N-acetylglucosamine moieties. (Mezgrhani, H. et al. (1997) J. Biol. Chem. 272:23340-23346.)

Regulation of levels of circulating thyroid hormone and glycosylated immunoglobulins is important for proper embryonic and fetal development and maintains metabolic activity. Thyroid hormones regulate metabolic rate. They increase metabolism by increasing the rate of carbohydrate absorption in the intestine, regulating lipid metabolism, and increasing mitochondrial respiration and oxidative phosphorylation. Thyroid hormones also influence body growth and nervous system development in the fetus. Increases in the level of circulating IgG with terminal GlcNAc residues appears to be associated with rheumatoid arthritis, Crohn's disease, and Sjogren's syndrome. (Bond, A. et al. (1997) J. Autoimmun. 10:77-85.)

Olfactomedin Family Proteins

The glycoprotein olfactomedin is a major component of the extracellular mucous matrix secreted by olfactory epithelial cells (Snyder, D. A. et al. (1991) Biochemistry 30:9143-9153; Yokoe, H. and Anholt, R. R. H. (1993) Proc. Natl. Acad. Sci. USA 90:4655-4659). In vivo, olfactomedin exists as a homodimer, the subunits of which are joined by intermolecular disulfide bonds and carbohydrate interactions. Olfactomedin may contribute to the protective and lubricating functions of mucous and/or promote the interaction of odorants and olfactory receptors in the nasal tract.

The recent identification of several proteins with similarity to olfactomedin suggests that olfactomedin and related proteins may define a distinct family of extracellular matrix glycoproteins. Two glycoprotein isoforms, designated neuronal olfactomedin-related glycoprotein-AMZ (NORG-AMZ) and -BMZ (NORG-BMZ), have been identified in rat brain (Danielson, P. E. et al. (1994) J. Neurosci. Res. 38:468-478). The carboxy-terminal 175 amino acids of both NORG-AMZ and NORG-BMZ share 33% amino acid identity and 19% conservative amino acid substitutions with the C-terminal region of olfactomedin. The mRNAs encoding NORG-AMZ and NORG-BMZ are transcribed from alternative promoters of a single gene and differ in their 5' ends. In situ hybridization and Northern analysis indicate that the mRNAs are expressed exclusively in the brain, particularly in neurons of the cortex and the hippocampus. Both NORG-AMZ and NORG-BMZ are predicted to contain signal peptides at their N-termini, suggesting that these proteins are both secreted. In addition, homologs of these proteins have also been discovered in mice which are virtually identical to the rat proteins (Nagano, T. et al. (1998) Brain Res. Mol. Brain Res. 53:13-23).

Another olfactomedin-related protein has been implicated in the development of primary open angle glaucoma (POAG) (Fingert, J. H. et al. (1998) Genome Res. 8:377-384). POAG is the most prevalent form of glaucoma, a disease which may cause irreversible blindness if left untreated. In the United States alone, glaucoma affects about 2.3 million people and blinds about 12,000 per year. POAG is characterized by progressive degeneration and cupping of the optic nerve, loss of peripheral visual field, and increased intraocular pressure. The development of both acquired and inherited forms of POAG is correlated with the presence of mutations in the *GLC1A* gene. Non-disease-causing polymorphisms have also been identified in *GLC1A*. *GLC1A* has been cloned from humans and mice and encodes a protein called myocilin. The C-terminal 250 amino acids of human myocilin shares 35% identity with the C-terminal region of olfactomedin. *GLC1A* is expressed in the eye, particularly in the ciliary body, retina, iris, and trabecular meshwork. *GLC1A* is also expressed in a variety of other tissues including the heart, skeletal muscle, stomach, thyroid, trachea, bone marrow, thymus, prostate, small intestine, and colon. *GLC1A* expression was not detected in brain, placenta, liver, kidney, spleen, or leukocytes.

Lectins

Lectins comprise a ubiquitous family of extracellular glycoproteins which bind cell surface carbohydrates specifically and reversibly, resulting in the agglutination of cells. (Reviewed in Drickamer, K. and Taylor, M. E. (1993) Annu. Rev. Cell Biol. 9:237-264.) This function is particularly important for activation of the immune response. Lectins mediate the agglutination and mitogenic stimulation of lymphocytes at sites of inflammation (Lasky, L. A. (1991) J. Cell. Biochem. 45:139-146; Paietta, E. et al. (1989) J. Immunol. 143:2850-2857).

Animal lectins have been grouped into four distinct families: 1) C-type lectins, which include selectins; 2) P-type lectins; 3) galectins (formerly termed S-type lectins or S-Lac lectins); and 4) pentraxins (Barondes, S.H. et al. (1994) J. Biol. Chem. 269:20807-20810). The C-type lectins bind carbohydrate ligands in a Ca^{2+}-dependent manner and are structurally related to the asialoglycoprotein receptor. Selectins, a subcategory of the C-type lectins, are composite transmembrane molecules which are involved in cell-cell interactions. The selectins include lymphocyte homing receptors and platelet/endothelial cell surface receptors (Stoolman, L.M. (1989) Cell 56:907-910).

C-type lectins contain Ca^{2+}-dependent carbohydrate-recognition domains. While prototypical C-type lectins are integral membrane proteins (e.g., the asialoglycoprotein receptor), a number of soluble C-type lectins have been identified. Examples of C-type lectins include pulmonary surfactant proteins A and D (SP-A and SP-D), mammalian C1q protein, bovine collectin-43, bovine conglutinin, mouse intelectin, and mannose-binding protein.

SP-D is a member of a family of host defense lectins, designated collectins. Collectins function in innate immunity in the lung and circulatory system to protect against a wide variety of potential pathogens, including viruses, bacteria, and fungi (Reid, K.B. (1998) Immunobiology, 199:200-207). Collectins have a collagen-like N-terminus domain which includes a number of Gly-Xaa-Xaa repeat,; an internal region termed the neck region, and a carbohydrate-binding C-terminus domain. SP-D is synthesized only in the lung (Lim et al. (1993) Immunology, 78:159-165). SP-D inhibits lectin- and anti-CD3-stimulated proliferation of human leukocytes in vitro, and this inhibition is associated with decreased cellular production of interleukin-2. (Borron, P.J. et al. (1998) J. Immunol. 161:4599-4603.)

Mouse intelectin is a homolog of the frog oocyte lectin and is expressed in the murine intestitial paneth cells of the small intestine. Frog oocyte lectin participates in the formation of the fertilization envelope and prevents polyspermy. Komiya et al. (1998, Biochem. Biophys. Res. Comm. 251:759-762) suggest that murine intelectin may be involved in host defense against microorganisms.

The galectin subfamily includes lectins that bind β-galactoside carbohydrate moieties in a thiol-dependent manner. (Reviewed in Hadari, Y. R. et al. (1995) J. Biol. Chem. 270:3447-3453.) Galectins are widely expressed and developmentally regulated. Because all galectins lack an N-terminal signal peptide, it is suggested that galectins are externalized through an atypical secretory mechanism. Two classes of galectins have been defined based on molecular weight and oligomerization properties. Small galectins form homodimers and are about 14 to 16 kilodaltons in mass, while large galectins are monomeric and about 29-37 kilodaltons. Galectins contain a characteristic carbohydrate recognition domain (CRD), also known as a galaptin domain. The CRD is about 140 amino acids long and contains several conserved residues. Secondary structure predictions indicate that the CRD forms several β-sheets. Known galectins contain one or two CRDs. (See Prosite PDOC00279 Vertebrate galactoside-binding lectin signature.)

Galectins play a number of roles in diseases and conditions associated with cell-cell and cell-matrix interactions. For example, certain galectins associate with sites of inflammation and bind to cell surface immunoglobulin E molecules. In addition, galectins may play an important role in cancer metastasis. Galectin overexpression is correlated with the metastatic potential of cancers in humans and mice. Moreover, anti-galectin antibodies inhibit processes associated with cell transformation, such as cell aggregation and anchorage-independent growth.

Galectin-8, also known as prostate carcinoma tumor antigen 1 (PCTA-1), is a novel galectin implicated in cancer progression (Su, Z.-Z. et al. (1996) Proc. Natl. Acad. Sci. USA 93:7252-7257). Galectin-8 was initially identified as the cell surface antigen recognized by a prostate tumor-directed monoclonal antibody, Pro 1.5. The galectin-8 gene encodes a 317-amino acid protein which contains two CRDs. Galectin-8 is expressed in invasive prostate carcinomas and early-stage prostate cancers, but not in normal prostate or benign prostatic hypertrophic tissue. In addition, galectin-8 is shed from the surface of cultured prostate cancer cells into the growth media. Together, these results suggest that detection of galectin-8 may be useful for early diagnosis of prostate cancer and that levels of galectin-8 in the circulation may correlate with disease progression. In addition, preliminary studies in mice suggest that the monoclonal antibody Pro 1.5 may itself be an effective therapeutic agent against tumor progression.

The discovery of new human carbohydrate-associated proteins and the polynucleotides encoding them satisfies a need in the art by providing new compositions which are useful in the diagnosis, prevention, and treatment of autoimmune/ inflammatory disorders, gastrointestinal disorders, infectious disorders, reproductive disorders, neurological disorders, eye disorders, and cell proliferative disorders, including cancer.

SUMMARY OF THE INVENTION

The invention is based on the discovery of new human carbohydrate-associated proteins, referred to collectively as "CRBAP" and individually as "CRBAP-1," "CRBAP-2," "CRBAP-3," "CRBAP-4," "CRBAP-5," "CRBAP-6," and "CRBAP-7". In one aspect, the invention provides a substantially purified polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-7, and fragments thereof.

The invention further provides a substantially purified variant having at least 90% amino acid sequence identity to the amino acid sequence of SEQ ID NO:1-7, and fragments thereof. The invention also provides an isolated and purified polynucleotide encoding the polypeptide comprising the amino acid sequence of SEQ ID NO:1-7, and fragments thereof. The invention also includes an isolated and purified polynucleotide variant having at least 70% polynucleotide sequence identity to the polynucleotide encoding the polypeptide comprising the amino acid sequence of SEQ ID NO:1-7, and fragments thereof.

The invention further provides an isolated and purified polynucleotide which hybridizes under stringent conditions to the polynucleotide encoding the polypeptide comprising the amino acid sequence of SEQ ID NO:1-7, or fragments thereof, as well as an isolated and purified polynucleotide having a sequence which is complementary to the polynucleotide encoding the polypeptide comprising the amino acid sequence of SEQ ID NO:1-7, and fragments thereof.

The invention also provides a method for detecting a polynucleotide in a sample containing nucleic acids, the method comprising the steps of (a) hybridizing the complement of the polynucleotide sequence to at least one of the polynucleotides of the sample, thereby forming a hybridization complex; and (b) detecting the hybridization complex, wherein the presence of the hybridization complex correlates with the presence of a polynucleotide in the sample. In one aspect, the method further comprises amplifying the polynucleotide prior to hybridization.

The invention also provides an isolated and purified polynucleotide comprising the polynucleotide sequence of SEQ ID NO:8-14, and fragments thereof. The invention further provides an isolated and purified polynucleotide comprising the polynucleotide sequence of SEQ ID NO:8 which contains a single nucleotide polymorphism with C replacing T at position 428. The invention further provides an isolated and purified polynucleotide variant having at least 70% polynucleotide sequence identity to the polynucleotide comprising the polynucleotide sequence of SEQ ID NO:8-14, and fragments thereof. The invention also provides an isolated and purified polynucleotide having a sequence complementary to the polynucleotide comprising the polynucleotide sequence of SEQ ID NO:8-14, and fragments thereof.

The invention further provides an expression vector containing at least a fragment of the

polynucleotide encoding the polypeptide comprising the sequence of SEQ ID NO:1-7, and fragments thereof. In another aspect, the expression vector is contained within a host cell.

The invention also provides a method for producing a polypeptide, the method comprising the steps of: (a) culturing the host cell containing an expression vector containing at least a fragment of a polynucleotide under conditions suitable for the expression of the polypeptide; and (b) recovering the polypeptide from the host cell culture.

The invention also provides a pharmaceutical composition comprising a substantially purified polypeptide having the sequence of SEQ ID NO:1-7, or fragments thereof in conjunction with a suitable pharmaceutical carrier.

The invention further includes a purified antibody which binds to a polypeptide comprising the sequence of SEQ ID NO:1-7, and fragments thereof, as well as a purified agonist and a purified antagonist of the polypeptide.

The invention also provides a method for treating or preventing a disorder associated with decreased expression or activity of CRBAP, the method comprising administering to a subject in need of such treatment an effective amount of a pharmaceutical composition comprising a substantially purified polypeptide having the amino acid sequence of SEQ ID NO:1-7, and fragments thereof, in conjunction with a suitable pharmaceutical carrier.

The invention also provides a method for treating or preventing a disorder associated with increased expression or activity of CRBAP, the method comprising administering to a subject in need of such treatment an effective amount of an antagonist of the polypeptide having the amino acid sequence of SEQ ID NO:1-7, and fragments thereof.

BRIEF DESCRIPTION OF THE FIGURES AND TABLES

Figures 1A and 1B show the amino acid sequence alignment between CRBAP-1 (714029; SEQ ID NO:1) and human galectin-8 (GI 2810994; SEQ ID NO:15), produced using the multisequence alignment program of LASERGENE software (DNASTAR, Madison WI).

Figures 2A, 2B, 2C, 2D, and 2E show the amino acid sequence alignment between CRBAP-2 (1450775; SEQ ID NO:2) and Clostridium perfringens hyaluronidase (GI 144861; SEQ ID NO:16).

Figures 3A and 3 B show the amino acid sequence alignment between CRBAP-3 (3369350; SEQ ID NO:3) and human TGR-CL10C (GI 1247124; SEQ ID NO:17), produced using the multisequence alignment program of LASERGENE software (DNASTAR, Madison WI).

Figures 4A and 4B show the amino acid sequence alignment among residues 147 through 402 of CRBAP-4 (1648214; SEQ ID NO:4), residues 159 through 409 of CRBAP-5 (2743295; SEQ

ID NO:5), and residues 204 through 457 of rat NORG-AMZ (GI 442368; SEQ ID NO:18), produced using the multisequence alignment program of LASERGENE software (DNASTAR, Madison WI).

Figures 5A and 5B show the amino acid sequence alignment between CRBAP-6 (2821011; SEQ ID NO:6) and bovine pulmonary SP-D (GI 415939; SEQ ID NO:19), produced using the multisequence alignment program of LASERGENE software (DNASTAR, Madison WI).

Figures 6A and 6B show the amino acid sequence alignment between CRBAP-7 (2921920; SEQ ID NO:7) and murine intelectin (GI 3357909; SEQ ID NO:20), produced using the multisequence alignment program of LASERGENE software.

Table 1 shows nucleotide and polypeptide sequence identification numbers (SEQ ID NO), clone identification numbers (clone ID), cDNA libraries, and cDNA fragments used to assemble full-length sequences encoding CRBAP.

Table 2 shows features of each polypeptide sequence including potential motifs, homologous sequences, and methods and algorithms used for identification of CRBAP.

Table 3 shows the tissue-specific expression patterns of each nucleic acid sequence as determined by northern analysis, diseases or disorders associated with these tissues, and the vector into which each cDNA was cloned.

Table 4 describes the tissues used to construct the cDNA libraries from which Incyte cDNA clones encoding CRBAP were isolated.

Table 5 shows the tools, programs, and algorithms used to analyze CRBAP along with applicable descriptions, references, and threshold parameters.

DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular machines, materials and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention

belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

DEFINITIONS

"CRBAP" refers to the amino acid sequences of substantially purified CRBAP obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and preferably the human species, from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist" refers to a molecule which, when bound to CRBAP, increases or prolongs the duration of the effect of CRBAP. Agonists may include proteins, nucleic acids, carbohydrates, or any other molecules which bind to and modulate the effect of CRBAP.

An "allelic variant" is an alternative form of the gene encoding CRBAP. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. Any given natural or recombinant gene may have none, one, or many allelic forms. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

"Altered" nucleic acid sequences encoding CRBAP include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as CRBAP or a polypeptide with at least one functional characteristic of CRBAP. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding CRBAP, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding CRBAP. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent CRBAP. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of CRBAP is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, positively charged

amino acids may include lysine and arginine, and amino acids with uncharged polar head groups having similar hydrophilicity values may include leucine, isoleucine, and valine; glycine and alanine; asparagine and glutamine; serine and threonine; and phenylalanine and tyrosine.

The terms "amino acid" or "amino acid sequence" refer to an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. In this context, "fragments," "immunogenic fragments," or "antigenic fragments" refer to fragments of CRBAP which are preferably at least 5 to about 15 amino acids in length, most preferably at least 14 amino acids, and which retain some biological activity or immunological activity of CRBAP. Where "amino acid sequence" is recited to refer to an amino acid sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification" relates to the production of additional copies of a nucleic acid sequence. Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known in the art.

The term "antagonist" refers to a molecule which, when bound to CRBAP, decreases the amount or the duration of the effect of the biological or immunological activity of CRBAP. Antagonists may include proteins, nucleic acids, carbohydrates, antibodies, or any other molecules which decrease the effect of CRBAP.

The term "antibody" refers to intact molecules as well as to fragments thereof, such as Fab, $F(ab')_2$, and Fv fragments, which are capable of binding the epitopic determinant. Antibodies that bind CRBAP polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant" refers to that fragment of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (given regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "antisense" refers to any composition containing a nucleic acid sequence which is

complementary to the "sense" strand of a specific nucleic acid sequence. Antisense molecules may be produced by any method including synthesis or transcription. Once introduced into a cell, the complementary nucleotides combine with natural sequences produced by the cell to form duplexes and to block either transcription or translation. The designation "negative" can refer to the antisense strand, and the designation "positive" can refer to the sense strand.

The term "biologically active," refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" refers to the capability of the natural, recombinant, or synthetic CRBAP, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

The terms "complementary" or "complementarity" refer to the natural binding of polynucleotides by base pairing. For example, the sequence "5' A-G-T 3'" bonds to the complementary sequence "3' T-C-A 5'." Complementarity between two single-stranded molecules may be "partial," such that only some of the nucleic acids bind, or it may be "complete," such that total complementarity exists between the single stranded molecules. The degree of complementarity between nucleic acid strands has significant effects on the efficiency and strength of the hybridization between the nucleic acid strands. This is of particular importance in amplification reactions, which depend upon binding between nucleic acids strands, and in the design and use of peptide nucleic acid (PNA) molecules.

A "composition comprising a given polynucleotide sequence" or a "composition comprising a given amino acid sequence" refer broadly to any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotide sequences encoding CRBAP or fragments of CRBAP may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

"Consensus sequence" refers to a nucleic acid sequence which has been resequenced to resolve uncalled bases, extended using XL-PCR kit (Perkin-Elmer, Norwalk CT) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from the overlapping sequences of more than one Incyte Clone using a computer program for fragment assembly, such as the GELVIEW Fragment Assembly system (GCG, Madison WI). Some sequences have been both extended and assembled to produce the consensus sequence.

The term "correlates with expression of a polynucleotide" indicates that the detection of the

presence of nucleic acids, the same or related to a nucleic acid sequence encoding CRBAP, by northern analysis is indicative of the presence of nucleic acids encoding CRBAP in a sample, and thereby correlates with expression of the transcript from the polynucleotide encoding CRBAP.

A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative" refers to the chemical modification of a polypeptide sequence, or a polynucleotide sequence. Chemical modifications of a polynucleotide sequence can include, for example, replacement of hydrogen by an alkyl, acyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

The term "similarity" refers to a degree of complementarity. There may be partial similarity or complete similarity. The word "identity" may substitute for the word "similarity." A partially complementary sequence that at least partially inhibits an identical sequence from hybridizing to a target nucleic acid is referred to as "substantially similar." The inhibition of hybridization of the completely complementary sequence to the target sequence may be examined using a hybridization assay (Southern or northern blot, solution hybridization, and the like) under conditions of reduced stringency. A substantially similar sequence or hybridization probe will compete for and inhibit the binding of a completely similar (identical) sequence to the target sequence under conditions of reduced stringency. This is not to say that conditions of reduced stringency are such that non-specific binding is permitted, as reduced stringency conditions require that the binding of two sequences to one another be a specific (i.e., a selective) interaction. The absence of non-specific binding may be tested by the use of a second target sequence which lacks even a partial degree of complementarity (e.g., less than about 30% similarity or identity). In the absence of non-specific binding, the substantially similar sequence or probe will not hybridize to the second non-complementary target sequence.

The phrases "percent identity" or "% identity" refer to the percentage of sequence similarity found in a comparison of two or more amino acid or nucleic acid sequences. Percent identity can be determined electronically, e.g., by using the MEGALIGN program (DNASTAR) which creates alignments between two or more sequences according to methods selected by the user, e.g., the clustal method. (See, e.g., Higgins, D.G. and P.M. Sharp (1988) Gene 73:237-244.) The clustal algorithm groups sequences into clusters by examining the distances between all pairs. The clusters are aligned pairwise and then in groups. The percentage similarity between two amino acid

sequences, e.g., sequence A and sequence B, is calculated by dividing the length of sequence A, minus the number of gap residues in sequence A, minus the number of gap residues in sequence B, into the sum of the residue matches between sequence A and sequence B, times one hundred. Gaps of low or of no similarity between the two amino acid sequences are not included in determining percentage similarity. Percent identity between nucleic acid sequences can also be counted or calculated by other methods known in the art, e.g., the Jotun Hein method. (See, e.g., Hein, J. (1990) Methods Enzymol. 183:626-645.) Identity between sequences can also be determined by other methods known in the art, e.g., by varying hybridization conditions.

"Human artificial chromosomes" (HACs) are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size, and which contain all of the elements required for stable mitotic chromosome segregation and maintenance.

The term "humanized antibody" refers to antibody molecules in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

"Hybridization" refers to any process by which a strand of nucleic acid binds with a complementary strand through base pairing.

The term "hybridization complex" refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., C_0t or R_0t analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words "insertion" or "addition" refer to changes in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively, to the sequence found in the naturally occurring molecule.

"Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

The term "microarray" refers to an arrangement of distinct polynucleotides on a substrate.

The terms "element" or "array element" in a microarray context, refer to hybridizable polynucleotides arranged on the surface of a substrate.

The term "modulate" refers to a change in the activity of CRBAP. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other

biological, functional, or immunological properties of CRBAP.

The phrases "nucleic acid" or "nucleic acid sequence," as used herein, refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material. In this context, "fragments" refers to those nucleic acid sequences which comprise a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:8-14, for example, as distinct from any other sequence in the same genome. For example, a fragment of SEQ ID NO:8-14 is useful in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:8-14 from related polynucleotide sequences. A fragment of SEQ ID NO:8-14 is at least about 15-20 nucleotides in length. The precise length of the fragment of SEQ ID NO:8-14 and the region of SEQ ID NO:8-14 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment. In some cases, a fragment, when translated, would produce polypeptides retaining some functional characteristic, e.g., antigenicity, or structural domain characteristic, e.g., ATP-binding site, of the full-length polypeptide.

The terms "operably associated" or "operably linked" refer to functionally related nucleic acid sequences. A promoter is operably associated or operably linked with a coding sequence if the promoter controls the translation of the encoded polypeptide. While operably associated or operably linked nucleic acid sequences can be contiguous and in the same reading frame, certain genetic elements, e.g., repressor genes, are not contiguously linked to the sequence encoding the polypeptide but still bind to operator sequences that control expression of the polypeptide.

The term "oligonucleotide" refers to a nucleic acid sequence of at least about 6 nucleotides to 60 nucleotides, preferably about 15 to 30 nucleotides, and most preferably about 20 to 25 nucleotides, which can be used in PCR amplification or in a hybridization assay or microarray. "Oligonucleotide" is substantially equivalent to the terms "amplimer," "primer," "oligomer," and "probe," as these terms are commonly defined in the art.

"Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

The term "sample" is used in its broadest sense. A sample suspected of containing nucleic acids encoding CRBAP, or fragments thereof, or CRBAP itself, may comprise a bodily fluid; an

extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

The terms "specific binding" or "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, or an antagonist. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide containing the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

The term "stringent conditions" refers to conditions which permit hybridization between polynucleotides and the claimed polynucleotides. Stringent conditions can be defined by salt concentration, the concentration of organic solvent, e.g., formamide, temperature, and other conditions well known in the art. In particular, stringency can be increased by reducing the concentration of salt, increasing the concentration of formamide, or raising the hybridization temperature.

The term "substantially purified" refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least about 60% free, preferably about 75% free, and most preferably about 90% free from other components with which they are naturally associated.

A "substitution" refers to the replacement of one or more amino acids or nucleotides by different amino acids or nucleotides, respectively.

"Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

"Transformation" describes a process by which exogenous DNA enters and changes a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed" cells includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A "variant" of CRBAP polypeptides refers to an amino acid sequence that is altered by one or more amino acid residues. The variant may have "conservative" changes, wherein a substituted amino acid has similar structural or chemical properties (e.g., replacement of leucine with isoleucine). More rarely, a variant may have "nonconservative" changes (e.g., replacement of glycine with tryptophan). Analogous minor variations may also include amino acid deletions or insertions, or both. Guidance in determining which amino acid residues may be substituted, inserted, or deleted without abolishing biological or immunological activity may be found using computer programs well known in the art, for example, LASERGENE software (DNASTAR).

The term "variant," when used in the context of a polynucleotide sequence, may encompass a polynucleotide sequence related to CRBAP. This definition may also include, for example, "allelic" (as defined above), "splice," "species," or "polymorphic" variants. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternate splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or an absence of domains. Species variants are polynucleotide sequences that vary from one species to another. The resulting polypeptides generally will have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

THE INVENTION

The invention is based on the discovery of new human carbohydrate-associated proteins (CRBAP), the polynucleotides encoding CRBAP, and the use of these compositions for the diagnosis, treatment, or prevention of autoimmune/ inflammatory disorders, gastrointestinal disorders, infectious disorders, reproductive disorders, neurological disorders, eye disorders, and cell proliferative disorders, including cancer.

Table 1 lists the Incyte Clones used to derive full length nucleotide sequences encoding CRBAP. Columns 1 and 2 show the sequence identification numbers (SEQ ID NO) of the amino acid and nucleic acid sequences, respectively. Column 3 shows the clone ID of the Incyte Clone in which nucleic acids encoding each CRBAP were identified, and column 4 shows the cDNA libraries from which these clones were isolated. Column 5 shows Incyte clones, corresponding cDNA libraries, and shotgun sequences which are useful as fragments in hybridization technologies and are part of the consensus nucleotide sequence of each CRBAP.

The polynucleotide sequence encoding CRBAP-1 (SEQ ID NO:8) contains a single

nucleotide polymorphism with C replacing T at position 428. C was found at position 428 in 10% of the clones, while T was found at position 428 in 90% of the clones.

The columns of Table 2 show various properties of the polypeptides of the invention: column 1 references the SEQ ID NO; column 2 shows the number of amino acid residues in each polypeptide; column 3, potential phosphorylation sites; column 4, potential glycosylation sites; column 5, the amino acid residues comprising signature sequences and motifs; column 6, the identity of each protein; and column 7, analytical methods used to identify each protein through sequence homology and protein motifs.

PFAM analysis indicates that CRBAP-1 resembles vertebrate galactoside-binding lectins from G58 through D165. As shown in Figures 1A and 1B, CRBAP-1 has chemical and structural similarity with human galectin-8 (GI 2810994; SEQ ID NO:15). In particular, CRBAP-1 and human galectin-8 share 28% identity. CRBAP-1 and the first galaptin domain of human galectin-8 (residues 1 through 153) share 23% identity. CRBAP-1 and the second galaptin domain of human galectin-8 (residues 185 through 315) share 34% identity.

As shown in Figures 2A, 2B, 2C, 2D, and 2E, CRBAP-2 has chemical and structural similarity with Clostridium perfringens hyaluronidase (GI 144861; SEQ ID NO:16). In particular, CRBAP-2 and Clostridium perfringens hyaluronidase share 14% identity.

As shown in Figures 3A and 3B, CRBAP-3 has chemical and structural similarity with human TGR-CL10C (GI 1247124; SEQ ID NO:17). In particular, CRBAP-3 and human TGR-CL10C share 30% identity over the length of CRBAP-3. In addition, CRBAP-3 and TGR-CL10C share 100% identity from residue P128 through H220.

As shown in Figures 4A and 4B, the C-terminus of CRBAP-4 from G147 to K402 shares 25% amino acid sequence identity with the C-terminus of rat NORG-AMZ (GI 442368; SEQ ID NO:18) from G204 to L457. In particular, the potential phosphorylation sites at T268, T295, and S352 and the predicted signal peptide in CRBAP-4 are conserved in rat NORG-AMZ. In addition, the C-terminus of CRBAP-4 from G147 to K402 also shares 24% amino acid sequence identity with the C-terminus of myocilin (GI 3065679; sequence not shown). As also shown in Figures 4A and 4B, the C-terminus of CRBAP-5 from T159 to S409 shares 33% amino acid sequence identity with the C-terminus of rat NORG-AMZ from G204 to L457. In particular, the potential phosphorylation sites at S215, S277, and T381 in CRBAP-5 are conserved in rat NORG-AMZ.

As shown in Figures 5A and 5B, CRBAP-6 has chemical and structural similarity with bovine lung SP-D (GI 415939; SEQ ID NO:19). In particular, CRBAP-6 and bovine lung SP-D share 32% identity, the signal peptide sequence, the C-type lectin and the C1q domain protein sequences, four conserved cysteine residues at C170, C242, C256, and C264, and a collagen-like

domain.

As shown in Figures 6A and 6B, CRBAP-7 has chemical and structural similarity with murine intelectin (GI 3357909; SEQ ID NO:20). In particular, CRBAP-7 and murine intelectin share 79% identity, five potential casein kinase II phosphorylation sites, one potential protein kinase C phosphorylation site, and have similar isoelectric points of 8.1 and 7.8, respectively.

The columns of Table 3 show the tissue-specificity and disease-association of nucleotide sequences encoding CRBAP. The first column of Table 3 lists the polynucleotide sequence identifiers. The second column lists tissue categories which express CRBAP as a fraction of total tissue categories expressing CRBAP. The third column lists the disease classes associated with those tissues expressing CRBAP. The fourth column lists the vectors used to subclone the cDNA library. Of particular note is the expression of CRBAP-1 and CRBAP-2 in libraries associated with cancer and cell proliferation, inflammation and immune response, reproductive tissues, and nervous tissues. Also of particular note is the expression of CRBAP-3 in reproductive, nervous, gastrointestinal tissues, the expression of CRBAP-6 in liver, kidney, ovary, gut, adrenal gland, and secretory epithelium, and the expression of CRBAP-7 in small intestine with Crohn's disease, and with lung, ovary, testis, and secretory epithelium.

The invention also encompasses CRBAP variants. A preferred CRBAP variant is one which has at least about 80%, more preferably at least about 90%, and most preferably at least about 95% amino acid sequence identity to the CRBAP amino acid sequence, and which contains at least one functional or structural characteristic of CRBAP.

The invention also encompasses polynucleotides which encode CRBAP. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising the sequence of SEQ ID NO:8-14, which encodes CRBAP.

The invention also encompasses a variant of a polynucleotide sequence encoding CRBAP. In particular, such a variant polynucleotide sequence will have at least about 70%, more preferably at least about 85%, and most preferably at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding CRBAP. A particular aspect of the invention encompasses a variant of SEQ ID NO:8-14 which has at least about 70%, more preferably at least about 85%, and most preferably at least about 95% polynucleotide sequence identity to SEQ ID NO:8-14. Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of CRBAP.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding CRBAP, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be

produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring CRBAP, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode CRBAP and its variants are preferably capable of hybridizing to the nucleotide sequence of the naturally occurring CRBAP under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding CRBAP or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding CRBAP and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of DNA sequences which encode CRBAP and CRBAP derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding CRBAP or any fragment thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:8-14, or fragments thereof, under various conditions of stringency. (See, e.g., Wahl, G.M. and S.L. Berger (1987) Methods Enzymol. 152:399-407; Kimmel, A.R. (1987) Methods Enzymol. 152:507-511.) For example, stringent salt concentration will ordinarily be less than about 750 mM NaCl and 75 mM trisodium citrate, preferably less than about 500 mM NaCl and 50 mM trisodium citrate, and most preferably less than about 250 mM NaCl and 25 mM trisodium citrate. Low stringency hybridization can be obtained in the absence of organic solvent, e.g., formamide, while high stringency hybridization can be obtained in the presence of at least about 35% formamide, and most preferably at least about 50% formamide. Stringent temperature conditions will ordinarily include temperatures of at least about 30°C, more preferably of at least about 37°C, and most preferably of at least about 42°C. Varying additional parameters, such as hybridization time, the concentration of detergent, e.g., sodium dodecyl sulfate (SDS), and the inclusion or exclusion of carrier DNA, are well known to those skilled in the art. Various levels of stringency are

accomplished by combining these various conditions as needed. In a preferred embodiment, hybridization will occur at 30°C in 750 mM NaCl, 75 mM trisodium citrate, and 1% SDS. In a more preferred embodiment, hybridization will occur at 37°C in 500 mM NaCl, 50 mM trisodium citrate, 1% SDS, 35% formamide, and 100 µg/ml denatured salmon sperm DNA (ssDNA). In a most preferred embodiment, hybridization will occur at 42°C in 250 mM NaCl, 25 mM trisodium citrate, 1% SDS, 50 % formamide, and 200 µg/ml ssDNA. Useful variations on these conditions will be readily apparent to those skilled in the art.

The washing steps which follow hybridization can also vary in stringency. Wash stringency conditions can be defined by salt concentration and by temperature. As above, wash stringency can be increased by decreasing salt concentration or by increasing temperature. For example, stringent salt concentration for the wash steps will preferably be less than about 30 mM NaCl and 3 mM trisodium citrate, and most preferably less than about 15 mM NaCl and 1.5 mM trisodium citrate. Stringent temperature conditions for the wash steps will ordinarily include temperature of at least about 25°C, more preferably of at least about 42°C, and most preferably of at least about 68°C. In a preferred embodiment, wash steps will occur at 25°C in 30 mM NaCl, 3 mM trisodium citrate, and 0.1% SDS. In a more preferred embodiment, wash steps will occur at 42°C in 15 mM NaCl, 1.5 mM trisodium citrate, and 0.1% SDS. In a most preferred embodiment, wash steps will occur at 68°C in 15 mM NaCl, 1.5 mM trisodium citrate, and 0.1% SDS. Additional variations on these conditions will be readily apparent to those skilled in the art.

Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Perkin-Elmer), thermostable T7 polymerase (Amersham Pharmacia Biotech, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Life Technologies, Gaithersburg MD). Preferably, sequence preparation is automated with machines such as the Robbins Hydra microdispenser (Robbins Scientific, Sunnyvale CA), Hamilton MICROLAB 2200 (Hamilton, Reno NV), Peltier Thermal Cycler 200 (PTC200; MJ Research, Watertown MA) and the ABI CATALYST 800 (Perkin-Elmer). Sequencing is then carried out using the ABI 373 or 377 DNA sequencing systems (Perkin-Elmer), the MEGABACE 1000 DNA sequencing system (Molecular Dynamics, Sunnyvale CA), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art. (See, e.g., Ausubel, F.M. (1997) Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853.)

The nucleic acid sequences encoding CRBAP may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector. (See, e.g., Sarkar, G. (1993) PCR Methods Applic. 2:318-322.) Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences. (See, e.g., Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186.) A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119.) In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-306). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 primer analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

When screening for full-length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Perkin-Elmer), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments

which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode CRBAP may be cloned in recombinant DNA molecules that direct expression of CRBAP, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced and used to express CRBAP.

The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter CRBAP-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

In another embodiment, sequences encoding CRBAP may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) Nucl. Acids Res. Symp. Ser. 215-223, and Horn, T. et al. (1980) Nucl. Acids Res. Symp. Ser. 225-232.) Alternatively, CRBAP itself or a fragment thereof may be synthesized using chemical methods. For example, peptide synthesis can be performed using various solid-phase techniques. (See, e.g., Roberge, J.Y. et al. (1995) Science 269:202-204.) Automated synthesis may be achieved using the ABI 431A peptide synthesizer (Perkin-Elmer). Additionally, the amino acid sequence of CRBAP, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography. (See, e.g, Chiez, R.M. and F.Z. Regnier (1990) Methods Enzymol. 182:392-421.) The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (See, e.g., Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY.)

In order to express a biologically active CRBAP, the nucleotide sequences encoding CRBAP or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotide sequences encoding CRBAP. Such elements may vary in their strength and

specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences encoding CRBAP. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where sequences encoding CRBAP and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used. (See, e.g., Scharf, D. et al. (1994) Results Probl. Cell Differ. 20:125-162.)

Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding CRBAP and appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. (See, e.g., Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel, F.M. et al. (1995) Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, ch. 9, 13, and 16.)

A variety of expression vector/host systems may be utilized to contain and express sequences encoding CRBAP. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotide sequences encoding CRBAP. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding CRBAP can be achieved using a multifunctional E. coli vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1 plasmid (Life Technologies). Ligation of sequences encoding CRBAP into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for in vitro transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence. (See, e.g., Van Heeke, G. and S.M. Schuster

(1989) J. Biol. Chem. 264:5503-5509.) When large quantities of CRBAP are needed, e.g. for the production of antibodies, vectors which direct high level expression of CRBAP may be used. For example, vectors containing the strong, inducible T5 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of CRBAP. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast Saccharomyces cerevisiae or Pichia pastoris. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign sequences into the host genome for stable propagation. (See, e.g., Ausubel, 1995, supra; Grant et al. (1987) Methods Enzymol. 153:516-54; and Scorer, C. A. et al. (1994) Bio/Technology 12:181-184.)

Plant systems may also be used for expression of CRBAP. Transcription of sequences encoding CRBAP may be driven by viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) EMBO J. 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (See, e.g., Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; and Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105.) These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. (See, e.g., The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196.)

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding CRBAP may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses CRBAP in host cells. (See, e.g., Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes. (See, e.g., Harrington, J.J. et al. (1997) Nat Genet. 15:345-355.)

For long term production of recombinant proteins in mammalian systems, stable expression of CRBAP in cell lines is preferred. For example, sequences encoding CRBAP can be transformed

into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk*⁻ or *apr*⁻ cells, respectively. (See, e.g., Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823.) Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* or *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g., Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14.) Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. 85:8047-8051.) Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech), ß glucuronidase and its substrate ß-glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding CRBAP is inserted within a marker gene sequence, transformed cells containing sequences encoding CRBAP can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding CRBAP under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

In general, host cells that contain the nucleic acid sequence encoding CRBAP and that express CRBAP may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein

sequences.

Immunological methods for detecting and measuring the expression of CRBAP using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on CRBAP is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art. (See, e.g., Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St Paul MN, Sect. IV; Coligan, J. E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY; and Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ).

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding CRBAP include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding CRBAP, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Pharmacia Biotech, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding CRBAP may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode CRBAP may be designed to contain signal sequences which direct secretion of CRBAP through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different

host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38), are available from the American Type Culture Collection (ATCC, Bethesda MD) and may be chosen to ensure the correct modification and processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding CRBAP may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric CRBAP protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of CRBAP activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the CRBAP encoding sequence and the heterologous protein sequence, so that CRBAP may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel (1995, supra, ch 10). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In a further embodiment of the invention, synthesis of radiolabeled CRBAP may be achieved in vitro using the TNT rabbit reticulocyte lysate or wheat germ extract systems (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, preferably ^{35}S-methionine.

Fragments of CRBAP may be produced not only by recombinant production, but also by direct peptide synthesis using solid-phase techniques. (See, e.g., Creighton, supra pp. 55-60.) Protein synthesis may be performed by manual techniques or by automation. Automated synthesis may be achieved, for example, using the ABI 431A Peptide Synthesizer (Perkin-Elmer). Various fragments of CRBAP may be synthesized separately and then combined to produce the full length molecule.

THERAPEUTICS

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of CRBAP and human carbohydrate-associated proteins. In addition, the expression of CRBAP is closely associated with inflamed, gastrointestinal, infected, reproductive, neurological, proliferating, and cancerous tissues, tissues of the eye, and inflammation of the gut and secretory epithelium. Therefore, CRBAP appears to play a role in autoimmune/ inflammatory disorders, gastrointestinal disorders, infectious disorders, reproductive disorders, neurological disorders, eye disorders, and cell proliferative disorders, including cancer. In the treatment of neoplastic, autoimmune/inflammatory, gastrointestinal, and infectious disorders associated with increased CRBAP expression or activity, it is desirable to decrease the expression or activity of CRBAP. In the treatment of the above conditions associated with decreased CRBAP expression or activity, it is desirable to increase the expression or activity of CRBAP.

Therefore, in one embodiment, CRBAP or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of CRBAP. Examples of such disorders include, but are not limited to, an autoimmune/inflammatory disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyenodocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, X-linked agammaglobinemia of Bruton, common variable immunodeficiency (CVI), DiGeorge's syndrome (thymic hypoplasia), thymic dysplasia, isolated IgA deficiency, severe combined immunodeficiency disease (SCID), immunodeficiency with thrombocytopenia and eczema (Wiskott-Aldrich syndrome), Chediak-Higashi syndrome, chronic granulomatous diseases, hereditary angioneurotic edema, immunodeficiency associated with Cushing's disease, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a gastrointestinal disorder such as dysphagia, peptic esophagitis, esophageal spasm, esophageal stricture, esophageal carcinoma, dyspepsia, indigestion, gastritis, gastric carcinoma, anorexia, nausea, emesis,

gastroparesis, antral or pyloric edema, abdominal angina, pyrosis, gastroenteritis, intestinal obstruction, infections of the intestinal tract, peptic ulcer, cholelithiasis, cholecystitis, cholestasis, pancreatitis, pancreatic carcinoma, biliary tract disease, hepatitis, hyperbilirubinemia, cirrhosis, passive congestion of the liver, hepatoma, infectious colitis, ulcerative colitis, ulcerative proctitis, Crohn's disease, Whipple's disease, Mallory-Weiss syndrome, colonic carcinoma, colonic obstruction, irritable bowel syndrome, short bowel syndrome, diarrhea, constipation, gastrointestinal hemorrhage, and acquired immunodeficiency syndrome (AIDS) enteropathy; an infectious disorder such as a viral infection, e.g., those caused by adenoviruses (acute respiratory disease, pneumonia), arenaviruses (lymphocytic choriomeningitis), bunyaviruses (Hantavirus), coronaviruses (pneumonia, chronic bronchitis), hepadnaviruses (hepatitis), herpesviruses (herpes simplex virus, varicella-zoster virus, Epstein-Barr virus, cytomegalovirus), flaviviruses (yellow fever), orthomyxoviruses (influenza), papillomaviruses (cancer), paramyxoviruses (measles, mumps), picornoviruses (rhinovirus, poliovirus, coxsackie-virus), polyomaviruses (BK virus, JC virus), poxviruses (smallpox), reovirus (Colorado tick fever), retroviruses (human immunodeficiency virus, human T lymphotropic virus), rhabdoviruses (rabies), rotaviruses (gastroenteritis), and togaviruses (encephalitis, rubella), infections by bacterial agents classified as pneumococcus, staphylococcus, streptococcus, bacillus, corynebacterium, clostridium, meningococcus, gonococcus, listeria, moraxella, kingella, haemophilus, legionella, bordetella, gram-negative enterobacterium including shigella, salmonella, and campylobacter, pseudomonas, vibrio, brucella, francisella, yersinia, bartonella, norcardium, actinomyces, mycobacterium, spirochaetale, rickettsia, chlamydia, and mycoplasma, infections by fungal agents classified as aspergillus, blastomyces, dermatophytes, cryptococcus, coccidioides, malasezzia, histoplasma, and other fungal agents causing various mycoses, and infections by parasites classified as plasmodium or malaria-causing, parasitic entamoeba, leishmania, trypanosoma, toxoplasma, pneumocystis carinii, intestinal protozoa such as giardia, trichomonas, tissue nematodes such as trichinella, intestinal nematodes such as ascaris, lymphatic filarial nematodes, trematodes such as schistosoma, and cestodes such as tapeworm; a reproductive disorder such as disorders of prolactin production; infertility, including tubal disease, ovulatory defects, and endometriosis; disruptions of the estrous cycle, disruptions of the menstrual cycle, polycystic ovary syndrome, ovarian hyperstimulation syndrome, endometrial and ovarian tumors, uterine fibroids, autoimmune disorders, ectopic pregnancies, and teratogenesis; cancer of the breast, fibrocystic breast disease, and galactorrhea; disruptions of spermatogenesis, abnormal sperm physiology, cancer of the testis, cancer of the prostate, benign prostatic hyperplasia, prostatitis, Peyronie's disease, impotence, carcinoma of the male breast, and gynecomastia; a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's

disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease; prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome; fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis; inherited, metabolic, endocrine, and toxic myopathies; myasthenia gravis, periodic paralysis; mental disorders including mood, anxiety, and schizophrenic disorders; akathesia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, and Tourette's disorder; an eye disorder such as conjunctivitis, keratoconjunctivitis sicca, keratitis, episcleritis, iritis, posterior uveitis, glaucoma including POAG, amaurosis fugax, ischemic optic neuropathy, optic neuritis, Leber's hereditary optic neuropathy, toxic optic neuropathy, vitreous detachment, retinal detachment, cataract, macular degeneration, central serous chorioretinopathy, retinitis pigmentosa, melanoma of the choroid, retrobulbar tumor, and chiasmal tumor; and a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus.

In another embodiment, a vector capable of expressing CRBAP or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of CRBAP including, but not limited to, those described above.

In a further embodiment, a pharmaceutical composition comprising a substantially purified CRBAP in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of CRBAP including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of CRBAP may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of CRBAP including, but not limited to, those listed above.

In a further embodiment, an antagonist of CRBAP may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of CRBAP. Such disorders may include, but are not limited to, those autoimmune/ inflammatory disorders, gastrointestinal disorders, infectious disorders, reproductive disorders, neurological disorders, eye disorders, and cell proliferative disorders, including cancer, discussed above. In one aspect, an antibody which specifically binds CRBAP may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissue which express CRBAP.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding CRBAP may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of CRBAP including, but not limited to, those described above.

In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary sequences, or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of CRBAP may be produced using methods which are generally known in the art. In particular, purified CRBAP may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind CRBAP. Antibodies to CRBAP may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are especially preferred for therapeutic use.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and others may be immunized by injection with CRBAP or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are

especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to CRBAP have an amino acid sequence consisting of at least about 5 amino acids, and, more preferably, of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein and contain the entire amino acid sequence of a small, naturally occurring molecule. Short stretches of CRBAP amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to CRBAP may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. (See, e.g., Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. 80:2026-2030; and Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120.)

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) Proc. Natl. Acad. Sci. 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; and Takeda, S. et al. (1985) Nature 314:452-454.) Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce CRBAP-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., Burton D.R. (1991) Proc. Natl. Acad. Sci. 88:10134-10137.)

Antibodies may also be produced by inducing in vivo production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature. (See, e.g., Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. 86: 3833-3837; Winter, G. et al. (1991) Nature 349:293-299.)

Antibody fragments which contain specific binding sites for CRBAP may also be generated. For example, such fragments include, but are not limited to, F(ab')2 fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')2 fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al. (1989) Science 246:1275-1281.)

Various immunoassays may be used for screening to identify antibodies having the desired

specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between CRBAP and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering CRBAP epitopes is preferred, but a competitive binding assay may also be employed (Pound, supra).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for CRBAP. Affinity is expressed as an association constant, K_a, which is defined as the molar concentration of CRBAP-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The K_a determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple CRBAP epitopes, represents the average affinity, or avidity, of the antibodies for CRBAP. The K_a determined for a preparation of monoclonal antibodies, which are monospecific for a particular CRBAP epitope, represents a true measure of affinity. High-affinity antibody preparations with K_a ranging from about 10^9 to 10^{12} l/mole are preferred for use in immunoassays in which the CRBAP-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with K_a ranging from about 10^6 to 10^7 l/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of CRBAP, preferably in active form, from the antibody (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington DC; Liddell, J. E. and Cryer, A. (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is preferred for use in procedures requiring precipitation of CRBAP-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available. (See, e.g., Catty, supra, and Coligan et al. supra.)

In another embodiment of the invention, the polynucleotides encoding CRBAP, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, the complement of the polynucleotide encoding CRBAP may be used in situations in which it would be desirable to block the transcription of the mRNA. In particular, cells may be transformed with sequences complementary to polynucleotides encoding CRBAP. Thus, complementary molecules or fragments may be used to modulate CRBAP activity, or to achieve regulation of gene function. Such

technology is now well known in the art, and sense or antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding CRBAP.

Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. Methods which are well known to those skilled in the art can be used to construct vectors to express nucleic acid sequences complementary to the polynucleotides encoding CRBAP. (See, e.g., Sambrook, supra; Ausubel, 1995, supra.)

Genes encoding CRBAP can be turned off by transforming a cell or tissue with expression vectors which express high levels of a polynucleotide, or fragment thereof, encoding CRBAP. Such constructs may be used to introduce untranslatable sense or antisense sequences into a cell. Even in the absence of integration into the DNA, such vectors may continue to transcribe RNA molecules until they are disabled by endogenous nucleases. Transient expression may last for a month or more with a non-replicating vector, and may last even longer if appropriate replication elements are part of the vector system.

As mentioned above, modifications of gene expression can be obtained by designing complementary sequences or antisense molecules (DNA, RNA, or PNA) to the control, 5', or regulatory regions of the gene encoding CRBAP. Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, are preferred. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177.) A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding CRBAP.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides,

corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA sequences encoding CRBAP. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

Many methods for introducing vectors into cells or tissues are available and equally suitable for use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) Nature Biotechnology 15:462-466.)

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as dogs, cats, cows, horses, rabbits, monkeys, and most preferably, humans.

An additional embodiment of the invention relates to the administration of a pharmaceutical or sterile composition, in conjunction with a pharmaceutically acceptable carrier, for any of the therapeutic effects discussed above. Such pharmaceutical compositions may consist of CRBAP, antibodies to CRBAP, and mimetics, agonists, antagonists, or inhibitors of CRBAP. The compositions may be administered alone or in combination with at least one other agent, such as a

stabilizing compound, which may be administered in any sterile, biocompatible pharmaceutical carrier including, but not limited to, saline, buffered saline, dextrose, and water. The compositions may be administered to a patient alone, or in combination with other agents, drugs, or hormones.

The pharmaceutical compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

In addition to the active ingredients, these pharmaceutical compositions may contain suitable pharmaceutically-acceptable carriers comprising excipients and auxiliaries which facilitate processing of the active compounds into preparations which can be used pharmaceutically. Further details on techniques for formulation and administration may be found in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA).

Pharmaceutical compositions for oral administration can be formulated using pharmaceutically acceptable carriers well known in the art in dosages suitable for oral administration. Such carriers enable the pharmaceutical compositions to be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions, and the like, for ingestion by the patient.

Pharmaceutical preparations for oral use can be obtained through combining active compounds with solid excipient and processing the resultant mixture of granules (optionally, after grinding) to obtain tablets or dragee cores. Suitable auxiliaries can be added, if desired. Suitable excipients include carbohydrate or protein fillers, such as sugars, including lactose, sucrose, mannitol, and sorbitol; starch from corn, wheat, rice, potato, or other plants; cellulose, such as methyl cellulose, hydroxypropylmethyl-cellulose, or sodium carboxymethylcellulose; gums, including arabic and tragacanth; and proteins, such as gelatin and collagen. If desired, disintegrating or solubilizing agents may be added, such as the cross-linked polyvinyl pyrrolidone, agar, and alginic acid or a salt thereof, such as sodium alginate.

Dragee cores may be used in conjunction with suitable coatings, such as concentrated sugar solutions, which may also contain gum arabic, talc, polyvinylpyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures. Dyestuffs or pigments may be added to the tablets or dragee coatings for product identification or to characterize the quantity of active compound, i.e., dosage.

Pharmaceutical preparations which can be used orally include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a coating, such as glycerol or sorbitol. Push-fit capsules can contain active ingredients mixed with fillers or binders, such as lactose or

starches, lubricants, such as talc or magnesium stearate, and, optionally, stabilizers. In soft capsules, the active compounds may be dissolved or suspended in suitable liquids, such as fatty oils, liquid, or liquid polyethylene glycol with or without stabilizers.

Pharmaceutical formulations suitable for parenteral administration may be formulated in aqueous solutions, preferably in physiologically compatible buffers such as Hanks' solution, Ringer's solution, or physiologically buffered saline. Aqueous injection suspensions may contain substances which increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Additionally, suspensions of the active compounds may be prepared as appropriate oily injection suspensions. Suitable lipophilic solvents or vehicles include fatty oils, such as sesame oil, or synthetic fatty acid esters, such as ethyl oleate, triglycerides, or liposomes. Non-lipid polycationic amino polymers may also be used for delivery. Optionally, the suspension may also contain suitable stabilizers or agents to increase the solubility of the compounds and allow for the preparation of highly concentrated solutions.

For topical or nasal administration, penetrants appropriate to the particular barrier to be permeated are used in the formulation. Such penetrants are generally known in the art.

The pharmaceutical compositions of the present invention may be manufactured in a manner that is known in the art, e.g., by means of conventional mixing, dissolving, granulating, dragee-making, levigating, emulsifying, encapsulating, entrapping, or lyophilizing processes.

The pharmaceutical composition may be provided as a salt and can be formed with many acids, including but not limited to, hydrochloric, sulfuric, acetic, lactic, tartaric, malic, and succinic acid. Salts tend to be more soluble in aqueous or other protonic solvents than are the corresponding free base forms. In other cases, the preferred preparation may be a lyophilized powder which may contain any or all of the following: 1 mM to 50 mM histidine, 0.1% to 2% sucrose, and 2% to 7% mannitol, at a pH range of 4.5 to 5.5, that is combined with buffer prior to use.

After pharmaceutical compositions have been prepared, they can be placed in an appropriate container and labeled for treatment of an indicated condition. For administration of CRBAP, such labeling would include amount, frequency, and method of administration.

Pharmaceutical compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells or in animal models such as mice, rats, rabbits, dogs, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for

administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example CRBAP or fragments thereof, antibodies of CRBAP, and agonists, antagonists or inhibitors of CRBAP, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the ED_{50} (the dose therapeutically effective in 50% of the population) or LD_{50} (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, and it can be expressed as the LD_{50}/ED_{50} ratio. Pharmaceutical compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the ED_{50} with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting pharmaceutical compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1 μg to 100,000 μg, up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

DIAGNOSTICS

In another embodiment, antibodies which specifically bind CRBAP may be used for the diagnosis of disorders characterized by expression of CRBAP, or in assays to monitor patients being treated with CRBAP or agonists, antagonists, or inhibitors of CRBAP. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for CRBAP include methods which utilize the antibody and a label to detect CRBAP in human body fluids or in extracts of cells or tissues. The antibodies may be used with or

without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring CRBAP, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of CRBAP expression. Normal or standard values for CRBAP expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, preferably human, with antibody to CRBAP under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, preferably by photometric means. Quantities of CRBAP expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding CRBAP may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantitate gene expression in biopsied tissues in which expression of CRBAP may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of CRBAP, and to monitor regulation of CRBAP levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding CRBAP or closely related molecules may be used to identify nucleic acid sequences which encode CRBAP. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification (maximal, high, intermediate, or low), will determine whether the probe identifies only naturally occurring sequences encoding CRBAP, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and should preferably have at least 50% sequence identity to any of the CRBAP encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:8-14 or from genomic sequences including promoters, enhancers, and introns of the CRBAP gene.

Means for producing specific hybridization probes for DNAs encoding CRBAP include the cloning of polynucleotide sequences encoding CRBAP or CRBAP derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a

variety of reporter groups, for example, by radionuclides such as ^{32}P or ^{35}S, or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding CRBAP may be used for the diagnosis of disorders associated with expression of CRBAP. Examples of such disorders include, but are not limited to, an autoimmune/inflammatory disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyenodocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, X-linked agammaglobinemia of Bruton, common variable immunodeficiency (CVI), DiGeorge's syndrome (thymic hypoplasia), thymic dysplasia, isolated IgA deficiency, severe combined immunodeficiency disease (SCID), immunodeficiency with thrombocytopenia and eczema (Wiskott-Aldrich syndrome), Chediak-Higashi syndrome, chronic granulomatous diseases, hereditary angioneurotic edema, immunodeficiency associated with Cushing's disease, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a gastrointestinal disorder such as dysphagia, peptic esophagitis, esophageal spasm, esophageal stricture, esophageal carcinoma, dyspepsia, indigestion, gastritis, gastric carcinoma, anorexia, nausea, emesis, gastroparesis, antral or pyloric edema, abdominal angina, pyrosis, gastroenteritis, intestinal obstruction, infections of the intestinal tract, peptic ulcer, cholelithiasis, cholecystitis, cholestasis, pancreatitis, pancreatic carcinoma, biliary tract disease, hepatitis, hyperbilirubinemia, cirrhosis, passive congestion of the liver, hepatoma, infectious colitis, ulcerative colitis, ulcerative proctitis, Crohn's disease, Whipple's disease, Mallory-Weiss syndrome, colonic carcinoma, colonic obstruction, irritable bowel syndrome, short bowel syndrome, diarrhea, constipation, gastrointestinal hemorrhage, and acquired immunodeficiency syndrome (AIDS) enteropathy; an infectious disorder such as a viral infection, e.g., those caused by adenoviruses (acute respiratory disease, pneumonia), arenaviruses (lymphocytic choriomeningitis), bunyaviruses (Hantavirus), coronaviruses (pneumonia, chronic bronchitis), hepadnaviruses (hepatitis), herpesviruses (herpes simplex virus, varicella-zoster

virus, Epstein-Barr virus, cytomegalovirus), flaviviruses (yellow fever), orthomyxoviruses (influenza), papillomaviruses (cancer), paramyxoviruses (measles, mumps), picornoviruses (rhinovirus, poliovirus, coxsackie-virus), polyomaviruses (BK virus, JC virus), poxviruses (smallpox), reovirus (Colorado tick fever), retroviruses (human immunodeficiency virus, human T lymphotropic virus), rhabdoviruses (rabies), rotaviruses (gastroenteritis), and togaviruses (encephalitis, rubella), infections by bacterial agents classified as pneumococcus, staphylococcus, streptococcus, bacillus, corynebacterium, clostridium, meningococcus, gonococcus, listeria, moraxella, kingella, haemophilus, legionella, bordetella, gram-negative enterobacterium including shigella, salmonella, and campylobacter, pseudomonas, vibrio, brucella, francisella, yersinia, bartonella, norcardium, actinomyces, mycobacterium, spirochaetale, rickettsia, chlamydia, and mycoplasma, infections by fungal agents classified as aspergillus, blastomyces, dermatophytes, cryptococcus, coccidioides, malasezzia, histoplasma, and other fungal agents causing various mycoses, and infections by parasites classified as plasmodium or malaria-causing, parasitic entamoeba, leishmania, trypanosoma, toxoplasma, pneumocystis carinii, intestinal protozoa such as giardia, trichomonas, tissue nematodes such as trichinella, intestinal nematodes such as ascaris, lymphatic filarial nematodes, trematodes such as schistosoma, and cestodes such as tapeworm; a reproductive disorder such as disorders of prolactin production; infertility, including tubal disease, ovulatory defects, and endometriosis; disruptions of the estrous cycle, disruptions of the menstrual cycle, polycystic ovary syndrome, ovarian hyperstimulation syndrome, endometrial and ovarian tumors, uterine fibroids, autoimmune disorders, ectopic pregnancies, and teratogenesis; cancer of the breast, fibrocystic breast disease, and galactorrhea; disruptions of spermatogenesis, abnormal sperm physiology, cancer of the testis, cancer of the prostate, benign prostatic hyperplasia, prostatitis, Peyronie's disease, impotence, carcinoma of the male breast, and gynecomastia; a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease; prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome; fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial

nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis; inherited, metabolic, endocrine, and toxic myopathies; myasthenia gravis, periodic paralysis; mental disorders including mood, anxiety, and schizophrenic disorders; akathesia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, and Tourette's disorder; an eye disorder such as conjunctivitis, keratoconjunctivitis sicca, keratitis, episcleritis, iritis, posterior uveitis, glaucoma including POAG, amaurosis fugax, ischemic optic neuropathy, optic neuritis, Leber's hereditary optic neuropathy, toxic optic neuropathy, vitreous detachment, retinal detachment, cataract, macular degeneration, central serous chorioretinopathy, retinitis pigmentosa, melanoma of the choroid, retrobulbar tumor, and chiasmal tumor; and a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus. The polynucleotide sequences encoding CRBAP may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered CRBAP expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding CRBAP may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide sequences encoding CRBAP may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantitated and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of nucleotide sequences encoding CRBAP in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of CRBAP, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a

sequence, or a fragment thereof, encoding CRBAP, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of an abnormal amount of transcript (either under- or over-expressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding CRBAP may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced in vitro. Oligomers will preferably contain a fragment of a polynucleotide encoding CRBAP, or a fragment of a polynucleotide complementary to the polynucleotide encoding CRBAP, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or quantitation of closely related DNA or RNA sequences.

Methods which may also be used to quantitate the expression of CRBAP include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves. (See, e.g., Melby, P.C. et al. (1993) J. Immunol. Methods 159:235-244; Duplaa, C. et al. (1993) Anal. Biochem. 212:229-236.) The speed of quantitation of multiple samples may be accelerated by running the assay in an ELISA format where the oligomer of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as targets in a microarray. The microarray can be used to monitor the expression level of large numbers of genes simultaneously and to identify

genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, and to develop and monitor the activities of therapeutic agents.

Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. 94:2150-2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.)

In another embodiment of the invention, nucleic acid sequences encoding CRBAP may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries. (See, e.g., Harrington, J.J. et al. (1997) Nat Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; and Trask, B.J. (1991) Trends Genet. 7:149-154.)

Fluorescent in situ hybridization (FISH) may be correlated with other physical chromosome mapping techniques and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, supra, pp. 965-968.) Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) site. Correlation between the location of the gene encoding CRBAP on a physical chromosomal map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder. The nucleotide sequences of the invention may be used to detect differences in gene sequences among normal, carrier, and affected individuals.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the number or arm of a particular human chromosome is not known. New sequences can be assigned to chromosomal arms by physical mapping. This provides valuable information to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the disease or syndrome has been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988) Nature 336:577-580.) The nucleotide sequence of the subject invention may also be used to detect differences in the chromosomal location due to translocation, inversion,

etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, CRBAP, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between CRBAP and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with CRBAP, or fragments thereof, and washed. Bound CRBAP is then detected by methods well known in the art. Purified CRBAP can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding CRBAP specifically compete with a test compound for binding CRBAP. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with CRBAP.

In additional embodiments, the nucleotide sequences which encode CRBAP may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following preferred specific embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all patents, applications, and publications mentioned above and below, in particular U.S. Ser. No. [Attorney Docket No: PF-0604 P, filed October 1, 1998], U.S. Ser. No. [Attorney Docket No: PF-0605 P, filed October 6, 1998], U.S. Ser. No. [Attorney Docket No: PF-0639 P, filed November 13, 1998], and U.S. Ser. No. [Attorney Docket No: PF-0646 P, filed December 3, 1998] are hereby expressly incorporated by reference.

EXAMPLES

I. cDNA Library C nstructi n

RNA was purchased from Clontech or isolated from tissues described in Table 4. Some tissues were homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Life Technologies), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with chloroform. RNA was precipitated from the lysates with either isopropanol or sodium acetate and ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA purity. In some cases, RNA was treated with DNase. For most libraries, poly(A+) RNA was isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Valencia CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA purification kit (Ambion, Austin TX).

In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP vector system (Stratagene) or SUPERSCRIPT plasmid system (Life Technologies), using the recommended procedures or similar methods known in the art. (See, e.g., Ausubel, 1997, supra, units 5.1-6.6). Reverse transcription was initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Pharmacia Biotech) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), PSPORT1 plasmid (Life Technologies), or pINCY (Incyte Pharmaceuticals, Palo Alto CA). Recombinant plasmids were transformed into competent E. coli cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5α, DH10B, or ElectroMAX DH10B from Life Technologies.

II. Isolation of cDNA Clones

Plasmids were recovered from host cells by cell lysis. Plasmids were purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a

high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a Fluoroskan II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

III. Sequencing and Analysis

The cDNAs were prepared for sequencing using the ABI CATALYST 800 (Perkin-Elmer) or the HYDRA microdispenser (Robbins Scientific) or MICROLAB 2200 (Hamilton) systems in combination with the PTC-200 thermal cyclers (MJ Research). The cDNAs were sequenced using at least one of the following: the ABI PRISM 373 or 377 sequencing systems (Perkin-Elmer) and standard ABI protocols, base calling software, and kits; the MEGABACE 1000 DNA sequencing system (Molecular Dynamics); the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Perkin-Elmer); solutions and dyes from Amersham Pharmacia Biotech; or other methods known in the art. Reading frames for the ESTs were determined using standard methods (reviewed in Ausubel, 1997, supra, unit 7.7). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example V.

The polynucleotide sequences derived from cDNA and extension sequencing were assembled and analyzed using a combination of software programs which utilize algorithms well known to those skilled in the art. Table 5 summarizes the tools, programs, and algorithms used and provides applicable descriptions, references, and threshold parameters. The first column of Table 5 shows the tools, programs, and algorithms used, the second column provides brief descriptions, the third column presents appropriate references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score, the greater the homology between two sequences). Sequences were analyzed using MACDNASIS PRO software (Hitachi Software Engineering) and LASERGENE software (DNASTAR).

The polynucleotide sequences were validated by removing vector, linker, and polyA sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programming, and dinucleotide nearest neighbor analysis. The sequences were then queried against a selection of public databases, such as GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS to acquire annotation, using programs based on BLAST, FASTA, and BLIMPS. The sequences were assembled into full length polynucleotide sequences using programs based on Phred, Phrap, and Consed, and were screened for open reading

frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were translated to derive the corresponding full length amino acid sequences, and these full length sequences were subsequently analyzed by querying against databases such as the GenBank databases, SwissProt, BLOCKS, PRINTS, Prosite, and Hidden Markov Model (HMM)-based protein family databases such as PFAM. HMM is a probalistic approach which analyzes consensus primary structures of gene families. (See, e.g., Eddy, S.R. (1996) Cur. Opin. Str. Biol. 6:361-365.) Single nucleotide polymorphisms were discovered using Consed software (Gordon, D. et al. (1998) Genome Research 8:195-202; Nickerson, D.A. et al. (1997) Nucleic Acids Research 25:2745-2751).

The programs described above for the assembly and analysis of full length polynucleotide and amino acid sequences were used to identify polynucleotide sequence fragments from SEQ ID NO:8-14. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies were described in The Invention section above.

IV. Northern Analysis

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, supra, ch. 7; Ausubel, 1995, supra, ch. 4 and 16.)

Analogous computer techniques applying BLAST were used to search for identical or related molecules in nucleotide databases such as GenBank or LIFESEQ database (Incyte Pharmaceuticals, Palo Alto CA). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

$$\frac{\% \text{ sequence identity} \times \% \text{ maximum BLAST score}}{100}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. For example, with a product score of 40, the match will be exact within a 1% to 2% error, and, with a product score of 70, the match will be exact. Similar molecules are usually identified by selecting those which show product scores between 15 and 40, although lower scores may identify related molecules.

The results of northern analyses are reported as a percentage distribution of libraries in which the transcript encoding CRBAP occurred. Analysis involved the categorization of cDNA libraries by organ/tissue and disease. The organ/tissue categories included cardiovascular,

dermatologic, developmental, endocrine, gastrointestinal, hematopoietic/immune, musculoskeletal, nervous, reproductive, and urologic. The disease/condition categories included cancer, inflammation/trauma, cell proliferation, neurological, and pooled. For each category, the number of libraries expressing the sequence of interest was counted and divided by the total number of libraries across all categories. Percentage values of tissue-specific and disease- or condition-specific expression are reported in Table 3.

V. Extension of CRBAP Encoding Polynucleotides

The full length nucleic acid sequence of SEQ ID NO:8-14 were produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer, to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries were used to extend the sequence. If more than one extension was necessary or desired, additional or nested sets of primers were designed.

High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing Mg^{2+}, $(NH_4)_2SO_4$, and β-mercaptoethanol, Taq DNA polymerase (Amersham Pharmacia Biotech), ELONGASE enzyme (Life Technologies), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

The concentration of DNA in each well was determined by dispensing 100 µl PICO GREEN quantitation reagent (0.25% (v/v) PICO GREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5 µl of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5 µl to 10 µl aliquot of the reaction mixture was analyzed by electrophoresis on a 1 % agarose mini-gel to determine which reactions were successful in extending

the sequence.

The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Pharmacia Biotech). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Pharmacia Biotech), treated with *Pfu* DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent E. coli cells. Transformed cells were selected on antibiotic-containing media, individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Pharmacia Biotech) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries were reamplified using the same conditions as described above. Samples were diluted with 20% dimethysulphoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT kit (Amersham Pharmacia Biotech) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Perkin-Elmer).

In like manner, the nucleotide sequence of SEQ ID NO:8-14 is used to obtain 5' regulatory sequences using the procedure above, oligonucleotides designed for such extension, and an appropriate genomic library.

VI. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:8-14 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250 μCi of [^{32}P]-adenosine triphosphate (Amersham Pharmacia Biotech), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Pharmacia Biotech). An aliquot containing 10^7 counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases:

Ase I, Bgl II, Eco RI, Pst I, Xba1, or Pvu II (DuPont NEN).

The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under increasingly stringent conditions up to 0.1x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are analyzed and compared using autoradiography.

VII. Microarrays

A chemical coupling procedure and an ink jet device can be used to synthesize array elements on the surface of a substrate. (See, e.g., Baldeschweiler, supra.) An array analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced by hand or using available methods and machines and contain any appropriate number of elements. After hybridization, nonhybridized probes are removed and a scanner used to determine the levels and patterns of fluorescence. The degree of complementarity and the relative abundance of each probe which hybridizes to an element on the microarray may be assessed through analysis of the scanned images.

Full-length cDNAs, Expressed Sequence Tags (ESTs), or fragments thereof may comprise the elements of the microarray. Fragments suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). Full-length cDNAs, ESTs, or fragments thereof corresponding to one of the nucleotide sequences of the present invention, or selected at random from a cDNA library relevant to the present invention, are arranged on an appropriate substrate, e.g., a glass slide. The cDNA is fixed to the slide using, e.g., UV cross-linking followed by thermal and chemical treatments and subsequent drying. (See, e.g., Schena, M. et al. (1995) Science 270:467-470; Shalon, D. et al. (1996) Genome Res. 6:639-645.) Fluorescent probes are prepared and used for hybridization to the elements on the substrate. The substrate is analyzed by procedures described above.

VIII. Complementary Polynucleotides

Sequences complementary to the CRBAP-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring CRBAP. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of CRBAP. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a

complementary oligonucleotide is designed to prevent ribosomal binding to the CRBAP-encoding transcript.

IX. Expression f CRBAP

Expression and purification of CRBAP are achieved using bacterial or virus-based expression systems. For expression of CRBAP in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac* (*tac*) hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria express CRBAP upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of CRBAP in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant Autographica californica nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding CRBAP by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect Spodoptera frugiperda (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus. (See Engelhard, E. K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945.)

In most expression systems, CRBAP is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from Schistosoma japonicum, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Pharmacia Biotech). Following purification, the GST moiety can be proteolytically cleaved from CRBAP at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel (1995, supra, ch 10 and 16). Purified CRBAP obtained by these methods can be used directly in the following activity assay.

X. Demonstration of CRBAP Activity

CRBAP-1

CRBAP-1 activity is demonstrated as the ability to bind to β-galactoside sugars. CRBAP-1 is applied to a lactosyl-Sepharose column, and the column is eluted with 0.1 M lactose. The presence of CRBAP-1 in the eluate is detected by sodium dodecyl sulfate polyacrylamide gel electrophoresis and indicates the ability of CRBAP-1 to bind β-galactoside sugars.

CRBAP-2

CRBAP-2 activity is demonstrated as the ability to hydrolyze HA (Lepperdinger, supra). Radioactively labeled HA is immobilized on microtiter plates with the aid of 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide and N-hydroxy-sulfosuccinimide. The radioactivity solubilized after incubation with CRBAP-2 is measured using a liquid scintillation counter and is proportional to the CRBAP-2 in the starting sample.

CRBAP-3

CRBAP-3 activity is associated with its ability to recognize, bind, and transport N-acetylglucosamine-bearing proteins. Therefore, CRBAP-3 activity is measured as the level of binding of N-acetylglucosamine-bearing peptides using affinity chromatography. CRBAP-3 is attached to a solid phase (e.g. agarose) and ^{125}I-N3-labelled N-acetylglucosamine-containing peptide (e.g. bovine serum albumin; GlcNAc-BSA) is loaded onto the column and allowed to interact for 1 hour. The column is then washed with loading buffer and bound peptide is eluted by addition of buffer containing 0.5 M unlabelled N-acetylglucosamine. Fractions are collected and measured for radioactivity using a gamma counter. CRBAP-3 activity is directly proportional to the quantity of bound radioactive peptide.

CRBAP-4 and CRBAP-5

An assay for CRBAP-4 or CRBAP-5 is based on the biochemical properties of olfactomedin (Snyder, supra). Post-translational glycosylation results in the attachment of carbohydrate moieties to olfactomedin. Glycosylated olfactomedin monomers bind with high affinity to the lectin Ricinus communis agglutinin I (RCA), a plant protein that binds to carbohydrate, whereas other glycoproteins do not bind to RCA, suggesting that the carbohydrate composition of olfactomedin is unique. Therefore, the activity of CRBAP-4 or CRBAP-5 is demonstrated by its high-affinity binding to RCA. CRBAP-4 or CRBAP-5 is monomerized by treatment with reducing agent and further solubilized with CHAPS detergent. RCA conjugated to an inert resin is added to this solution. After an appropriate incubation period, the mixture is centrifuged to separate the RCA-resin and any material bound to it from the unbound fraction. The RCA-resin is washed with reducing agent in the presence of detergent. Material bound with high affinity is specifically eluted from the RCA-resin with D-galactose, a sugar that competes for RCA carbohydrate binding sites. The amount of CRBAP-4 or CRBAP-5 in the unbound fraction and the

eluent is analyzed by SDS-polyacrylamide gel electrophoresis and western blot utilizing specific antibody directed against CRBAP-4 or CRBAP-5. The ratio of CRBAP-4 or CRBAP-5 in the eluent to that in the unbound fraction is proportional to the amount of CRBAP-4 or CRBAP-5 bound to RCA and is a direct measure of CRBAP-4 or CRBAP-5 activity.

CRBAP-6 and CRBAP-7

CRBAP-6 or CRBAP-7 is measured by the ability of C-type lectin to bind carbohydrates. Carbohydrates may be demonstrated by examining the ability of recombinant CRBAP-6 or CRBAP-7 to bind to affinity columns comprising carbohydrates (e.g., lactose, maltose, D-mannose, D-galactose, which are available from Sigma Chemical Corp., St. Louis MO) or by using the assay described by Christa et al. (1994, FEBS Lett. 337:114-118). Preferably a soluble form of CRBAP-6 or CRBAP-7 (i.e., lacking the transmembrane domain) is expressed and applied to the affinity column.

Some C-type lectins are also known to agglutinate bacteria. The ability of CRBAP-6 or CRBAP-7 to agglutinate bacteria is examined using the assay described by Iovanna et al. (1991, J. Biol. Chem. 266:24664-24669). Briefly, bacteria (e.g., E. coli strains KH802 or JM101) are grown at 37°C to stationary phase in Luria-Bertani (LB) medium. The bacteria are then collected by centrifugation and washed in phosphate-buffered saline (PBS). The washed bacteria are resuspended in PBS containing 0.5 mM $CaCl_2$ (PBS/$CaCl_2$) and are placed in the wells of microtiter plates at a concentration of approximately 5 x 10^7 bacteria/200 µl PBS/$CaCl_2$. CRBAP-6 or CRBAP-7 (preferably in a soluble form) is then added at a variety of concentrations (e.g., 1 to 50 µg/ml) and the presence of macroscopic aggregation is monitored following a 3 hour incubation at 25°C. Concanavalin A and albumin at 50 µg/ml may be employed as positive and negative controls, respectively.

In another alternative, an assay for CRBAP-6 or CRBAP-7 activity measures the amount of cell aggregation induced by overexpression of CRBAP-6 or CRBAP-7. In this assay, cultured cells such as NIH 3T3 are transfected with cDNA encoding CRBAP-6 or CRBAP-7 contained within a suitable mammalian expression vector under control of a strong promoter. Cotransfection with cDNA encoding a fluorescent marker protein, such as Green Fluorescent Protein (Clontech), is useful for identifying stable transfectants. The amount of cell agglutination, or clumping, associated with transfected cells is compared with that associated with nontransfected cells. The amount of cell agglutination is proportional to CRBAP-6 or CRBAP-7 activity.

XI. Functional Assays

CRBAP function is assessed by expressing the sequences encoding CRBAP at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a

mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include pCMV SPORT (Life Technologies) and pCR3.1 (Invitrogen, Carlsbad CA), both of which contain the cytomegalovirus promoter. 5-10 µg of recombinant vector are transiently transfected into a human cell line, preferably of endothelial or hematopoietic origin, using either liposome formulations or electroporation. 1-2 µg of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP, and to evaluate cellular properties, for example, their apoptotic state. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M. G. (1994) Flow Cytometry, Oxford, New York NY.

The influence of CRBAP on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding CRBAP and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding CRBAP and other genes of interest can be analyzed by northern analysis or microarray techniques.

XII. Production of CRBAP Specific Antibodies

CRBAP substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) Methods Enzymol. 182:488-495), or other purification techniques, is used to immunize rabbits and to produce antibodies using standard protocols.

Alternatively, the CRBAP amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for

selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art. (See, e.g., Ausubel, 1995, supra, ch. 11.)

Typically, oligopeptides 15 residues in length are synthesized using an ABI 431A Peptide Synthesizer (Perkin-Elmer) using fmoc-chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity. (See, e.g., Ausubel, 1995, supra.) Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide activity by, for example, binding the peptide to plastic, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

XIII. Purification of Naturally Occurring CRBAP Using Specific Antibodies

Naturally occurring or recombinant CRBAP is substantially purified by immunoaffinity chromatography using antibodies specific for CRBAP. An immunoaffinity column is constructed by covalently coupling anti-CRBAP antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Pharmacia Biotech). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing CRBAP are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of CRBAP (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/CRBAP binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and CRBAP is collected.

XIV. Identification of Molecules Which Interact with CRBAP

CRBAP, or biologically active fragments thereof, are labeled with ^{125}I Bolton-Hunter reagent. (See, e.g., Bolton et al. (1973) Biochem. J. 133:529-539.) Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled CRBAP, washed, and any wells with labeled CRBAP complex are assayed. Data obtained using different concentrations of CRBAP are used to calculate values for the number, affinity, and association of CRBAP with the candidate molecules.

Various modifications and variations of the described methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

TABLE 1

Polypeptide SEQ ID NO:	Nucleotide SEQ ID NO:	Clone ID	Library	Fragments
1	8	714029	PROSTUT01	620990F1 (PGANNOT01), 714029H1, 714029X315F1, and 714029X328F1 (PROSTUT01), 1338783F1 (COLNTUT03), 1354746T6 (LUNGNOT09), 1376593F1 and 1376593T1 (LUNGNOT10), 2591130F6 (LUNGNOT22), 3427924T6 (BRSTNOR01), 3720846H1 (PENCNOT10), 4554976H1 (KERAUNT01), and SZAH02052F1
2	9	1450775	PENITUT01	1450775H1 (PENITUT01), 2197778F6 (SPLNFET02), 2216628H1 (SINTFET03), 2240704F6 (PANCTUT02), 2745765H1 (LUNGTUT11), 2830219F6 and 2830219H1 (TLYMNOT03), 2957078H1 (KIDNFET01), 3001459H1 (TLYMNOT06), and 3319368F6 (PROSBPT03)
3	10	3369350	CONNTUT04	1569729 (UTRSNOT05), 840896R1 (PROSTUT05), 3369350H1 (CONNTUT04), 1545419T1 (PROSTUT04), 864526R1 (BRAITUT03)
4	11	1648214	PROSTUT09	1648214H1 and 1648214X14 (PROSTUT09), 1258421F1 (MENITUT03), SAGA01200F1, SAGA02759F1, SBAA03843F2, SBAA01253F1, SAGA00437R1, SAGA02098F1
5	12	2743295	BRSTTUT14	2743295H1 and 2743295X310F1 (BRSTTUT14), 3109944H1 (BRSTTUT15), 1748824F6 (STOMTUT02), 4718554H1 (BRAIHCT02), 3820353H1 (BONSTUT01), 2598657F6 (UTRSNOT10), 960034R6 (BRSTTUT03), 1672040F6 (BLADNOT05)
6	13	2821011	ADRETUT06	2821011H1 and 2821011F6 (ADRETUT06). 4630524H1 (GBLADIT02), 2595195F6 and 2595945F6 (OVARTUT01), 988057T6 and 988057R6 (LVENNOT03)
7	14	2921920	SININOT04	2921920H1, 2921920F6, and 2921920T6 (SININOT04)

TABLE 2

Protein SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequence	Identification	Analytical Methods
1	171	S4 T72 S155 T92 T159	N20	Vertebrate Galactoside-binding lectin: G58 - D165	Galectin – 8 (g2810994) [Homo sapiens]	PFAM, BLAST, Motifs
2	666	T9 S112 T236 S263 T346 S364 S410 T452 T501 S505 S515 T533 T554 T153 S398 S601	N227 N424 N588 N633	Signal Peptide: M1 - G56	Hyaluronidase (g144861) [Clostridium perfrigens]	BLAST, SPScan, Motifs
3	307	S80 T194 S240 S210 T271		Signal Peptide: M1 – R37	TGR-CL10C (g1247124) [Homo sapiens]	BLAST, HMM, SPScan, Motifs
4	402	T118 T185 T232 T245 T268 T295 T350 S352 S150 S172 T390 T395 Y101 Y104	N66 N138 N183	Signal Peptide: M1 – A26 Potential glycosaminoglycan attachment site: S284 – G287 C-terminal region of extracellular proteins: G335 – P357	Neuronal olfactomedin-related ER localized protein (g442368) [Rattus norvegicus]	BLAST, BLOCKS, , HMM, SPScan, Motifs
5	409	S38 S97 S139 T197 S277 T317 T381 S97 T115 S215 T322 Y377	N354		Neuronal olfactomedin-related ER localized protein (g442368) [Rattus norvegicus]	BLAST, Motifs
6	271	S90 Y154		Signal Peptide: M1 – L20 Collagen-like domain: G41 – P112 C-type lectin domain: A247 – C256 C1q domain protein sequence: G44 – G77 C-type lectin family sequence: E150 – E265	Lung surfactant protein D (g415939) [Bos taurus]	BLAST, HMM, SPScan, BLOCKS, ProfileScan, PFAM
7	325	S30 S52 S62 T80 S96 T107 T142 S296 S52 S208		Signal Peptide: M1 – A26 Potential glycosaminoglycan attachment site: S86 – G89	Intelectin (g3357909) [Mus musculus]	BLAST, HMM, SPScan , Motifs

TABLE 3

SEQ ID NO:	Useful Fragment	Tissue Expression (Fraction of Total)	Disease or Condition (Fraction of Total)	Vector
8	465 - 500	Reproductive (0.327) Nervous (0.163) Gastrointestinal (0.122) Developmental (0.102)	Cell Proliferation (0.673) Inflammation (0.204)	PBLUESCRIPT
9	162 - 191	Reproductive (0.197) Hematopoietic/Immune (0.180) Cardiovascular (0.164) Nervous (0.148) Developmental (0.115)	Cell Proliferation (0.541) Inflammation (0.361)	pINCY
10	286 - 345	Reproductive (0.253) Nervous (0.200) Gastrointestinal (0.179)	Cancer (0.540) Inflammation (0.232)	pINCY
11	649 - 693	Reproductive (0.400) Gastrointestinal (0.160) Developmental (0.120) Nervous (0.120)	Cancer (0.800)	pINCY
12	334 - 378	Reproductive (0.480) Gastrointestinal (0.210)	Cancer (0.550) Inflammation (0.240)	pINCY
13	168 - 218	Gastrointestinal (0.368) Reproductive (0.158) Urologic (0.105) Cardiovascular (0.105) Developmental (0.105) Endocrine (0.105)	Cancer (0.368) Inflammation/trauma (0.315) Cell Proliferation (0.263)	pINCY
14	129 - 200	Gastrointestinal (0.556) Reproductive (0.222) Cardiovascular (0.222)	Cancer (0.556) Inflammation (0.444)	pINCY

TABLE 4

Nucleotide SEQ ID NO:	Clone ID	Library	Library Description
8	714029	PROSTUT01	Library was constructed using RNA isolated from the prostate tumor tissue removed from a 50-year-old Caucasian male. Pathology indicated grade 3 adenocarcinoma (Gleason grade 3+3). Patient history included dysuria, carcinoma in situ of prostate, coronary atherosclerosis, and hyperlipidemia.
9	1450775	PENITUT01	Library was constructed using RNA isolated from tumor tissue removed from the penis of a 64-year-old Caucasian male during penile amputation. Pathology indicated a fungating invasive grade 4 squamous cell carcinoma involving the inner wall of the foreskin and extending onto the glans penis. Patient history included benign neoplasm of the large bowel, atherosclerotic coronary artery disease, angina pectoris, gout, and obesity. Family history included malignant pharyngeal neoplasm, chronic lymphocytic leukemia, and chronic liver disease.
10	3369350	CONNTUT04	The library was constructed using RNA isolated from tumorous spinal soft tissue removed from a 35-year-old Caucasian male during an exploratory laparotomy. Pathology indicated schwannoma with degenerative changes. Patient history included anxiety, depression, benign hypertension, neurofibromatosis, and benign neoplasm of the scrotum. Family history included brain cancer, liver disease, and multiple sclerosis.
11	1648214	PROSTUT09	The library was constructed using RNA isolated from prostate tumor tissue removed from a 66-year-old Caucasian male during a radical prostatectomy, redical cystectomy, and urinary diversion. Pathology indicated grade 3 transitional cell carcinoma. Patient history included lung neoplasm, benign hypertension, and tobacco abuse in remission. Family history included malignant breast neoplasm, tuberculosis, benign hypertension, cerebrovascular disease, atherosclerotic coronary artery disease, and lung cancer.
12	2743295	BRSTTUT14	The library was constructed using RNA isolated from breast tumor tissue removed from a 62-year-old Caucasian female during a unilateral extended simple mastectomy. Pathology indicated an invasive grade 3 (of 4), nuclear grade 3 (of 3) adenocarcinoma, ductal type, located in the upper outer quadrant. Metastatic adenocarcinoma was identified in one (of 14) axillary lymph nodes. Immunohistochemical stains showed the tumor cells were strongly positive for estrogen receptors and weakly positive for progesterone receptors. Patient history included a benign colon neoplasm, hyperlipidemia, cardiac dysrhythmia, and obesity. Family history included atherosclerotic coronary artery disease, myocardial infarction, colon cancer, ovary cancer, lung cancer, and cerebrovascular disease.
13	2821011	ADRETUT06	The library was constructed using RNA isolated from adrenal tumor tissue removed from a 57-year-old Caucasian female during a unilateral right adrenalectomy. Pathology indicated pheochromocytoma (chromaffin cell tumor), forming a nodular mass completely replacing the medulla of the adrenal gland.
14	2921920	SININOT04	The library was constructed using RNA isolated from diseased ileum tissue removed from a 26-year-old Caucasian male during a partial colectomy, permanent colostomy, and an incidental appendectomy. Pathology indicated moderately to severely active Crohn's disease. Family history included regional enteritis of the small intestine.

Table 5

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Perkin-Elmer Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Perkin-Elmer Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Perkin-Elmer Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25: 3389-3402.	*ESTs:* Probability value= 1.0E-8 or less *Full Length sequences*: Probability value= 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises as least five functions: fasta, tfasta, fastx, tfastx, and ssearch.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad Sci. 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183: 63-98; and Smith, T.F. and M. S. Waterman (1981) Adv. Appl. Math. 2:482-489.	*ESTs:* fasta E value=1.06E-6 *Assembled ESTs*: fasta Identity= 95% or greater and Match length=200 bases or greater; fastx E value=1.0E-8 or less *Full Length sequences:* fastx score=100 or greater
BLIMPS	A BLocks IMProved Searcher that matches a sequence against those in BLOCKS and PRINTS databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S and J.G. Henikoff, Nucl. Acid Res., 19:6565-72, 1991. J.G. Henikoff and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Attwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37: 417-424.	Score=1000 or greater; Ratio of Score/Strength = 0.75 or larger; and Probability value= 1.0E-3 or less
PFAM	A Hidden Markov Models-based application useful for protein family search.	Krogh, A. et al. (1994) J. Mol. Biol., 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322.	Score=10-50 bits, depending on individual protein families

Table 5 (continued)

Program	Description	Reference	Parameter Threshold
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25: 217-221.	Score= 4.0 or greater
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M. S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M. S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score= 120 or greater; Match length= 56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12: 431-439.	Score=5 or greater
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch et al. supra; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	